MODIFICAÇÃO DE UM ALGORITMO PARA RESOLVER PROBLEMAS DE QUADRADOS MÍNIMOS NÃO LINEARES

RITA F. A. SANTOS

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

CAMPINAS - SÃO PAULO
BRASIL

MODIFICAÇÃO DE UM ALGORITMO PARA RESOLVER PROBLEMAS DE QUADRADOS MÍNIMOS NÃO LINEARES

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Srta. RITA FILOMENA ALVES DOS SANTOS e aprovada pela Comissão Julgadora.

Campinas, 20 de abril de 1989.

Prof.Dr. JOSÉ MARIO MARTÍNEZ
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de Mestrado em Matemática Aplicada.

AGRADECIMENTOS

A José Mário Martínez (Quase - Newton), pela orientação, atenção, disponibilidade, espontaneidade, compreensão e tolerância.

Aos Professores do Departamento de Matemática Aplicada pelo bom relacionamento, em especial a profa. Maria Aparecida (Cheti), que nos dispensou uma valiosa atenção.

Ao Dorival que nos prestou vários serviços de datilografia

Aos amigos do Mestrado em Matemática Aplicada pelo apoio e inúmeros cafezinhos pagos.

Aos meus amigos Renato e Amélia

À FAPESP pelo auxílio financeiro.

À D. Maria, "seu" Manoel

e ao Dira.

ÍNDICE

INTRODUÇÃO

Consideremos o problema de

Minimizar $\| F(x) \|_2^2$

com $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m \geq n$ e $F'(x)$ esparsa.

Estudaremos um algoritmo (ver [1]) para resolver este tipo de problema com as seguintes características:

(a) A "Equação Gauss-Newton" é resolvida parcialmente em cada passo usando um método de Gradientes Conjugados com um precondicionamento "ortogonal truncado".

(b) O novo ponto é obtido usando uma busca do tipo região de confiança bidimensional.

Neste trabalho, tentaremos melhorar o algoritmo supracitado em dois sentidos. Em primeiro lugar, melhorar o desempenho em problemas mal escalados, através da substituição de ∇F por $D\nabla F$, um gradiente convenientemente escalado. Em segundo lugar, simplificar o processo de busca , introduzindo uma busca parabólica, em vez do processo de região de confiança.

As etapas de nosso trabalho são as seguintes: no capítulo 1, daremos uma visão geral de alguns métodos para resolver o problema que estamos enfocando. No capítulo 2 apresentaremos o método de Levenberg-Marquardt, um dos métodos mais importantes empregados na resolução deste tipo de problema . No capítulo 3 faremos um estudo da teoria , e das diferentes partes do algoritmo supracitado. No capítulo 4 apresentaremos as modificações sugeridas para melhorar o algoritmo que é estudado

no capítulo 3. No capítulo 5 serão relatadas nossas experiências numéricas. E por último tentaremos tirar conclusões e fazer comentários sobre as experiências numéricas. Além disso colocaremos no apêndice A alguns teoremas auxiliares usados no texto principal.

CAPÍTULO I

QUADRADOS MÍNIMOS NÃO LINEARES

1.1. O Problema

O problema de quadrados mínimos não lineares consiste em

$$\text{Minimizar} \quad \frac{1}{2} \| F(x) \|_2^2 \qquad (1.1.1)$$

onde a função resíduo $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, com $m \geq n$, é não linear e $F_i(x)$ denota a i-ésima componente de $F(x)$. Tal problema surge frequentemente em ajuste de dados, onde é necessário escolher um x que minimize o resíduo $F_i(x) = m(x, t_i) - y_i$, com o objetivo de ajustar aos dados (t_i, y_i), $i=1,\ldots,m$, o modelo $m(x, t)$ que é não linear em x.

Se $m = n$, o problema se reduz à resolução de um sistema de equações não lineares.

O gradiente e a Hessiana da função objetivo (1.1.1), denotados respectivamente por $g(x)$ e $G(x)$, têm uma estrutura especial. Seja a matriz Jacobiana $m \times n$ de $F(x)$ denotada por $J(x)$ e seja a matriz $G_i(x)$ a matriz Hessiana de $F_i(x)$. Então

$$g(x) = J(x)^T F(x)$$

$$G(x) = J(x)^T J(x) + Q(x),$$

onde $Q(x) = \sum_{i=1}^{m} F_i(x)\ G_i(x)$. Observamos que a Hessiana da função objetivo consiste na combinação de derivadas parciais de primeira

e de segunda ordem.

Aplicando o método de Newton ao problema (1.1.1) obtemos

$$x_{k+1} = x_k - (J(x_k)^T J(x_k) + Q(x_k))^{-1} J(x_k)^T F(x_k) . \quad (1.1.2)$$

No decorrer deste capítulo discutiremos a aplicação deste método, dentre outros, ao problema de quadrados mínimos não lineares, distinguindo entre problemas de resíduo pequeno e resíduo grande.

1.2. Métodos do Tipo Gauss - Newton

Consideremos a aproximação afim de $F(x)$, na vizinhança de x_c:

$$m_c(x) = F(x_c) + J(x_c)(x - x_c) \quad (1.2.1)$$

onde $m_c(x)$: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ e $m \geq n$.

A melhor situação que poderíamos ter seria a de acharmos um x de forma que $m_c(x) = 0$, mas nem sempre isso é possível. Portanto, o caminho lógico para resolver o problema de quadrados mínimos não lineares é escolher um $\hat{x}$ como solução do problema de quadrados mínimos lineares

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \ \frac{1}{2} \| m_c(x) \|_2^2 . \quad (1.2.2)$$

Assumimos que $J(x_k)$ tem posto coluna completo. Assim, a solução de (1.2.2) é

$$\hat{x} = x_c - (J(x_c)^T J(x_c))^{-1} J(x_c)^T F(x_c). \qquad (1.2.3)$$

O método acima definido é chamado método de Gauss - Newton.

Comparando-o com o método de Newton para quadrados mínimos não lineares, observamos que as equações (1.1.2) e (1.2.3) diferem exatamente na matriz Hessiana $J(x)^T J(x) + Q(x)$, já que o termo $Q(x)$ é omitido no método de Gauss - Newton.

O método de Gauss - Newton está baseado na premissa de que eventualmente o termo de $1^{\underline{a}}$ ordem da Hessiana $J(x)^T J(x)$ dominará o termo de $2^{\underline{a}}$ ordem $Q(x)$, mas esta suposição só é válida quando o resíduo é zero na solução, ou é relativamente pequeno.

Este método tem convergência local quadrática, como o método de Newton, quando $F(x)$ é linear ou quando o resíduo é zero ($Q(x_*)=0$). Tem convergência local linear quando o resíduo é relativamente pequeno, ou seja, $Q(x_*)$ é pequeno em relaçao a $J(x_*)^T J(x_*)$. Não se verificará a convergência local quando o resíduo for grande , isto é, se $Q(x_*)$ for muito grande em relaçao a $J(x_*)^T J(x_*)$.

Confirmaremos o resultado acima pelo teorema e corolário seguintes:

TEOREMA 1.2.1: Sejam $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$, e $f(x) = \frac{1}{2} F(x)^T F(x)$, duas vezes diferenciáveis em um conjunto aberto convexo $D \subset \mathbb{R}^n$. Assumimos que $J(x)$ é Lipschitz, com $\| J(x) \|_2 \leq \alpha$ para todo $x \in D$, e que existem $x_* \in D$ e λ, $\rho \geq 0$ tais que $J(x_*)^T F(x_*) = 0$, λ é o menor autovalor de $J(x_*)^T J(x_*)$ e

$$\| (J(x) - J(x_*))^T F(x_*) \|_2 \leq \rho \| x - x_* \|_2 \qquad (1.2.4)$$

para todo $x \in D$. Se $\rho < \lambda$, então, para algum $c \in (1, \lambda/\rho)$, existe $\varepsilon > 0$ tal que para todo $x_0 \in N(x_*, c)$, a seqüencia gerada pelo método de Gauss - Newton

$$x_{k+1} = x_k - (J(x_k)^T J(x_k))^{-1} J(x_k)^T F(x_k)$$

é bem definida, converge para x_* e obedece a:

$$\| x_{k+1} - x_* \|_2 \leq \frac{c\rho}{\lambda} \| x_k - x_* \|_2 + \frac{c\alpha\gamma}{2\lambda} \| x_k - x_* \|_2^2 \qquad (1.2.5)$$

$$\| x_{k+1} - x_* \|_2 \leq \frac{c\rho + \lambda}{2\lambda} \| x_k - x_* \|_2 < \| x_k - x_* \|_2. \qquad (1.2.6)$$

Demonstração: Esta será feita por indução. Mostraremos primeiro que este teorema se verifica para k=0.

Assumimos que $\lambda > \rho \geq 0$ e seja $c \in (1, \lambda/\rho)$; chamaremos $J(x_0)$, $F(x_0)$ e $F(x_*)$ por J_0, F_0 e F_* respectivamente, assim como também substituíremos $\| \cdot \|_2$ por $\| \cdot \|$. Por um argumento familiar, existe $\varepsilon_1 > 0$ tal que $J_0^T J_0$ é não singular e

$$\| (J_0^T J_0)^{-1} \| \leq \frac{c}{\lambda} \quad p/\ x_0 \in N(x_*, \varepsilon_1). \qquad (1.2.7)$$

Seja

$$\varepsilon = \min \left\{ \varepsilon_1, \frac{\lambda - c\rho}{c\alpha\gamma} \right\}. \qquad (1.2.8)$$

claramente x_1 é bem definido, então

$$x_1 = x_0 - (J_0^T J_0)^{-1} J_0^T F_0$$

$$x_1 - x_* = x_0 - x_* - (J_0^T J_0)^{-1} J_0^T F_0$$

$$= -(J_0^T J_0)^{-1} \left[J_0^T F_0 + J_0^T J_0 (x_* - x_0) \right]$$

$$= -(J_0^T J_0)^{-1} \left[J_0^T F_* - J_0^T (F_* - F_0 - J_0(x_* - x_0)) \right] \qquad (1.2.9)$$

Pelo Lema 1 do apêndice A, temos,

$$\| F_* - F_0 - J_0(x_* - x_0) \| \leq \frac{\gamma}{2} \| x_0 - x_* \|^2. \qquad (1.2.10)$$

Relembrando que $J(x_*)^T F(x_*) = 0$, então, de (1.2.4), temos :

$$\| J_0^T F_* \| \leq \rho \| x_0 - x_* \|. \qquad (1.2.11)$$

Combinando (1.2.9), (1.2.7), (1.2.10) e (1.2.11) e relembrando que $\| J_0 \| \leq \alpha$, então:

$$\| x_1 - x_0 \| \leq \| (J_0^T J_0)^{-1} \| \left[\| J_0^T F_* \| + \| J_0 \| \| F_* - F_0 - J_0(x_* - x_0) \| \right]$$

$$\leq \frac{c}{\lambda} \left[\rho \| x_0 - x_* \| + \frac{\alpha\gamma}{2} \| x_0 - x \|^2 \right]$$

o que prova (1.2.5). Do resultado acima e de (1.2.8) temos

$$\| x_1 - x_* \| \leq \| x_0 - x_* \| \left[\frac{c\rho}{\lambda} + \frac{c\alpha\gamma}{2\lambda} \| x_0 - x_* \| \right]$$

$$\leq \| x_0 - x_* \| \left[\frac{c\rho}{\lambda} + \frac{\lambda - c\gamma}{2\lambda} \right]$$

$$= \frac{c\rho + 2\lambda}{2\lambda} \| x_0 - x_* \|$$

$$< \| x_0 - x_* \|.$$

o que prova (1.2.6).

Agora suponhamos que a tese é válida para k. Então:

$$\| x_{k+1} - x_* \|_2 \leq \frac{c\rho}{\lambda} \| x_k - x_* \|_2 + \frac{c\alpha\gamma}{2\lambda} \| x_k - x_* \|_2^2$$

$$\| x_{k+1} - x_* \|_2 \leq \frac{c\rho + \lambda}{2\lambda} \| x_k - x_* \|_2 < \| x_k - x_* \|_2.$$

Para provar para k + 1, o procedimento é idêntico àquele usado quando k = 0.■

COROLÁRIO 1.2.2.: Com as suposições do teorema (1.2.1) e se F(x) = 0, então existe $\varepsilon > 0$ tal que para todo $x_0 \in N(x_*, \varepsilon)$, a sequência gerada pelo método de Gauss - Newton é bem definida e converge quadraticamente para x_* .

Demonstração: Como $F(x_*) = 0$ e tomando $\rho = 0$ em (1.2.4), deduzimos de (1.2.6) que a sequência $\{ x_k \}$ converge para x_*, e de (1.2.5) que a velocidade da convergência é quadrática.■

A partir da equação (1.2.4) vemos que ρ é uma medida absoluta combinada da não linearidade de F e da grandeza do resíduo do problema; Se F(x) é linear ou $F(x_*) = 0$ isto implica que $\rho = 0$. Portanto, o teorema diz que a rapidez da convergência do método de Gauss - Newton decresce com a relativa não linearidade ou com a relativa grandeza do resíduo do problema.

Abaixo, mostraremos as vantagens e desvantagens do método de Gauss - Newton.

Vantagens:

1- Tem convergência local quadrática nos problemas de resíduo zero.

2- Tem convergência local linear rápida nos problemas que são quase lineares e têm resíduos razoavelmente pequenos.

3- Resolve o problema de quadrados mínimos lineares em uma iteração.

Desvantagens:

1- Tem convergência local linear lenta em problemas que são suficientemente não lineares ou têm resíduos razoavelmente grandes.

2- Não converge localmente em problemas que são muito não lineares ou que possuem resíduos muitos grandes.

3- Não é bem definido quando $J(x_k)$ não tem posto coluna completo.

4- Necessariamente não converge globalmente.

Para conseguirmos uma modificação do método de forma a obter convergência global, comecemos observando que a direção $-[J(x_k)^T J(x_k)]^{-1} J(x_k)^T F(x_k)$ é de descida quando está bem definida. Sugerimos dois caminhos para melhorar o algoritmo de Gauss - Newton: usarmos uma busca linear ou uma região de confiança. Estas duas sugestões levam a dois algoritmos que são usados na prática.

O algoritmo em que o método de Gauss - Newton é combinado com uma busca linear consiste simplesmente em fazer:

$$x_{k+1} = x_k + \lambda_k (J(x_k)^T J(x_k))^{-1} J(x_k)^T F(x_k) \qquad (1.2.12)$$

onde λ_k pode ser escolhido por exemplo por qualquer algoritmo de busca linear. A convergência desta variante pode ser muito lenta nos problemas em que o método de Gauss - Newton " sem busca " teve dificuldades. O algoritmo de Gauss - Newton " com busca " também não é bem definido se $J(x_k)$ não tem posto coluna completo.

A outra modificação do algoritmo de Gauss - Newton é a escolha de x_+ usando uma região de confiança: assim resolvemos

$$\begin{aligned} &\underset{x \in \mathbb{R}^n}{\text{Min}} \; \| F(x) + J(x)(x_+ - x) \|_2 \qquad (1.2.13) \\ &\text{suj. a/} \; \| x_+ - x \|_2 \leq \Delta \end{aligned}$$

Pelo Lema 2 do apêndice A, provamos que a solução (1.2.13) é

$$x_+ = x - (J(x)^T J(x) + \mu I)^{-1} J(x)^T F(x) \qquad (1.2.14)$$

onde $\mu = 0$ se $\Delta \geq \| (J(x)^T J(x))^{-1} J(x)^T F(x) \|$ e $\mu > 0$ se $\Delta = \| (J(x)^T J(x))^{-1} J(x)^T F(x) \|$. A formula (1.2.14) foi sugerida primeiramente por Levenberg em (1944) e depois por Marquardt em (1963) e é conhecida como método de Levenberg - Marquardt.

As propriedades de convergência deste método são similares às do método de Gauss - Newton e são dadas no teorema (1.2.3).

TEOREMA 1.2.3: Se assumimos que as condições do teorema (1.2.1) são satisfeitas e que a sequência $\{ \mu_k \}$ de números reais não negativos é limitada por b, e se $\rho < \lambda$, então, para algum $c \in (1, (\lambda + b)/(\rho + b))$, e para $\varepsilon > 0$ tal que para todo $x_0 \in N(x_*, \varepsilon)$, a sequência gerada pelo método de Levenberg - Marquardt

$$x_{k+1} = x_k - (J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T F(x_k)$$

é bem definida e obedece a:

$$\| x_{k+1} - x_* \|_2 \leq \frac{c(\rho + b)}{(\lambda + b)} \| x_k - x_* \|_2 + \frac{c\alpha\gamma}{2(\lambda + b)} \| x_k - x_* \|_2^2$$

$$\| x_{k+1} - x_* \|_2 \leq \frac{c(\rho + b)}{(\lambda + b)} \| x_k - x_* \|_2 < \| x_k - x_* \|_2.$$

Se $F(x) = 0$ e $\mu_k = 0$ $(\| J(x_k)^T F(x_k) \|_2)$, então $\{ x_k \}$ converge quadraticamente.

Demonstração: É uma extensão direta do teorema (1.2.1) e do corolário (1.2.2).■

O algoritmo de Levenberg - Marquardt é preferido em relação ao algoritmo de Gauss - Newton " com busca " por vários fatores. Por exemplo, o método de Levenberg - Marquardt é bem definido mesmo quando $J(x_k)$ não tem posto coluna completo. Por outro lado, quando o passo de Gauss - Newton é muito longo, o passo de Levenberg - Marquardt está mais perto de ser uma direção de máxima descida e é geralmente superior àquele.

1.3. Métodos Tipo Newton

A outra classe de métodos considerados para resolver problemas de quadrados mínimos não lineares é baseada no modelo quadrático da série de Taylor. Para problemas de quadrados mínimos não lineares o método de Newton consiste na escolha de um x_+ como ponto crítico deste modelo:

$$x_+ = x - (J(x)^T J(x) + Q(x))^{-1} J(x)^T F(x).$$

O método de Newton, apesar de possuir boas propriedades de convergência local, é raramente usado para este tipo de problema, pois o cálculo de $J(x)^T J(x) + Q(x)$ é normalmente muito caro.

Outros métodos considerados são os métodos secantes, mas uma aplicação direta destes também não é desejável, porque aproximam a $\nabla^2 f(x)$ de forma completa. (Ver [2])

1.4. Critérios de Parada

Se o menor valor possivel de $f(x) = \frac{1}{2}F(x)^T F(x)$ é zero , então o teste

$$f(x_+) \cong 0 \ ?$$

é adequado; mas isto é um critério apropriado de convergência somente para problemas de resíduo nulo. Portanto deve ser implementado como $f(x_+) \leq TOL$, onde a tolerância TOL é escolhida de forma apropriada.

O segundo teste é o da convergência pelo gradiente

$$\nabla f(x_+) = J(x_+)^T F(x_+) \cong 0.$$

Isto tem um sentido especial, visto que pode ser interpretado como a pergunta se $F(x_+)$ é quase ortogonal ao subespaço linear gerado pelas colunas de $J(x_+)$.

CAPÍTULO II

O MÉTODO DE LEVENBERG - MARQUARDT

2.1. Introdução

O método de Levenberg - Marquardt, que já foi citado no capítulo anterior, é uma variação do método de Gauss - Newton. Numa rápida visão deste método, observa-se que ele toma um passo reduzido para calcular um novo ponto. Suponhamos que primeiro seja escolhido o comprimento do passo e depois seja usado um modelo quadrático n-dimensional para escolher a direção. Em outras palavras, suponhamos que temos um x e alguma estimativa Δ para o comprimento máximo do passo. A questão é: Qual é o melhor passo dentro do comprimento máximo ? Respondendo esta questão, diremos que o melhor passo é a solução de:

$$\begin{aligned} &\text{Min } q(x + P) = f(x) + \nabla f(x)^T P + \frac{1}{2} P^T \nabla f^2(x) P \\ &\text{suj. a } \| P \|_2 \leq \Delta \, . \end{aligned} \qquad (2.1.1)$$

O problema (2.1.1) é a base do modelo de " Região de Confiança ".

2.2 O Método

Consideremos a solução de:

$$\underset{P \in \mathbb{R}^n}{\text{Minimize}} \ \frac{1}{2} \| J_k P + F_k \|_2^2 \qquad (2.2.1)$$

$$\text{suj. a} \ \| P \|_2 \leq \Delta$$

para algum Δ.

Pelas condições de Kuhn-Tucker se P é mínimo local do subproblema e os gradientes das restrições ativas ($\| P \|_2 = \Delta$) em P são linearmente independentes, temos então que

$$\nabla f(P) = -\mu_k \nabla(\| P \|_2^2 - \Delta)$$

ou seja

$$J_k^T J_k P + J_k^T F_k = \mu_k P$$

logo

$$(J_k^T J_k + \mu_k I)P = -J_k^T F_k \qquad (2.2.2)$$

onde μ_k é um escalar não negativo, o multiplicador de Lagrange.

A solução P de (2.2.2) é definida como direção de busca de Levenberg-Marquardt.

Um bom valor para μ_k (ou Δ) deve ser escolhido de forma que assegure o decréscimo suficiente. Existe uma relação entre os escalares μ_k e Δ de forma que temos que se $\mu_k = 0$, P_k torna-se direção de Gauss - Newton, isto é, $\| P \|_2 \leq \Delta$, para Δ bastante grande e quando $\mu \longrightarrow \infty$ $\| p \|_2 \longrightarrow 0$, P_k torna-se paralelo a direção de máxima descida. Isto implica que $F(x_k + P_k)$ é suficientemente menor que F_k para Δ bastante pequeno. Confirmaremos o resultado acima, mostrando a relação entre os escalares μ_k e Δ.

Por (2.2.2) temos que

$$P_k = -(J_k^T J_k + \mu_k I)^{-1} J_k^T F_k, . \qquad (2.2.3)$$

Normalizando a expressão (2.2.3),

$$\| P_k \|_2 = \| -(J_k^T J_k + \mu_k I)^{-1} J_k^T F_k \|_2 .$$

Assim, pelo teorema 3 do apêndice A, temos que

$$\| P_k \|_2 \leq \| (J_k^T J_k + \mu_k I)^{-1} \|_F \| J_k^T F_k \|_2$$

$$\Longrightarrow \qquad \| P_k \|_2 \leq n^{1/2} \| (J_k^T J_k + \mu_k I)^{-1} \|_2 \| J_k^T F_k \|_2$$

como a matriz $(J_k^T J_k + \mu_k I)^{-1}$ é simétrica e pelos teoremas 4 e 5 do apêndice A,

$$\| P_k \|_2 \leq n^{1/2} \frac{1}{|\rho_{min} + \mu_k|} \beta$$

onde ρ_{min} é o menor autovalor de $(J_k^T J_k)$ e $\beta = \| J_k^T F_k \|_2$.

Então se

$$\mu_k \longrightarrow \infty \Longrightarrow n^{1/2} \frac{1}{|\rho_{min} + \mu_k|} \beta \longrightarrow 0$$

ou

$$\mu_k = 0 \qquad \Longrightarrow n^{1/2} \frac{1}{|\rho_{min} + \mu_k|} \beta \geq 0.$$

Logo se tomamos $\Delta = n^{1/2} \frac{1}{|\rho_{min} + \mu_k|} \beta$, teremos que

$$\mu_k \longrightarrow \infty \Longrightarrow \Delta \longrightarrow 0 \text{ e } \| P \| \longrightarrow 0$$

ou

$$\mu_k = 0 \quad \Longrightarrow \Delta \geq 0 \text{ e } \| P \| \leq \Delta.$$

Seja $P_{LM}(\mu_k)$ uma solução de (2.2.2) (direção de Levenberg - Marquardt) para um valor específico x_k, onde μ_k é positivo, e suponhamos que J_k não tenha posto completo; então temos:

$$\frac{\| P_N - P_{LM}(\mu_k) \|}{\| P_N \|} = O(1),$$

onde P_N é a direção de Newton, e este resultado acima independe de μ_k.

Para atualizarmos Δ, podemos utilizar interpolação cúbica.

2.3. O uso do Método Levenberg - Marquardt nos Problemas de Grande Porte

Quando n é muito grande, duas dificuldades são encontradas : o tempo de computação requerido, que pode ser grande e não justificar a resolução do problema; e, o que é mais crítico, pode ser preciso usar armazenamento auxiliar. No caso do método de Levenberg - Marquardt, a principal dificuldade é que em cada iteração é resolvido um problema n-dimensional.

No próximo capítulo apresentaremos um algoritmo no qual calculamos a direção d que nos dá o novo $x_{k+1} = x_k + d$ usando uma implementação do método de Levenberg-Marquardt. Além disto vamos poder observar que fazemos uma busca bidimensional e não n-dimensional como no método de Levenberg - Marquardt, o que torna o algoritmo 3.2 muito mais econômico e sem incovenientes no caso de aplicação a problemas de grande porte.

CAPÍTULO III

A ESTRATÉGIA BIDIMENSIONAL PARA QUADRADOS MÍNIMOS NÃO LINEARES

3.1. Introdução

Estudaremos agora um algoritmo que procura conservar propriedades de convergência rápida em problemas de quadrados mínimos não lineares de resíduos pequenos e um comportamento razoável nos casos gerais. (Ver [1])

Os métodos do tipo Levenberg - Marquardt têm essas propriedades e são os mais populares.

Este algoritmo tem as seguintes características:

a_ Em cada iteração, o problema de quadrados mínimos lineares

$$J(x_k)\ W_k \approx -F(x_k) \qquad (3.1.1)$$

é resolvido parcialmente usando gradientes conjugados, de maneira que o método pode ser pensado como do tipo " Inexact Gauss - Newton ", que é análogo aos métodos Inexact - Newton. (Ver [7,8])

b_ O sistema (3.1.1) é precondicionado por um método tipo Ortogonal Truncado. A mesma matriz de precondicionamento pode ser usada em várias iterações consecutivas.

c_ O ponto x_{k+1} é obtido por um processo de busca no plano gerado pelo gradiente de $\| F(x) \|_2^2$ em x_k e por W_k, solução de (3.1.1).

3.2. Princípios Básicos

Seja $F: D \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$, $m \geq n$, $F \in C^1(D)$, D aberto. Seja $x_0 \in D$ um ponto inicial arbitrário, (η_k) uma seqüencia de números positivos tal que $\lim \eta_k = 0$, $\theta_1, \theta_2 \in (0, 1)$, $\theta_3 \in (0, 1/2)$, e $0 < \underline{M} < \bar{M} < \infty$.

Dado x_k, a k-ésima aproximação da solução obtida com o algoritmo, denotamos $F_k = F(x_k)$, $J_k = J(x_k)$ (matriz jacobiana de F em x_k) e $\| \cdot \| = \| \cdot \|_2$.

Os passos para obter x_{k+1} são os seguintes:

ALGORITMO 3.2

Passo 1: Calcular $g_k = J_k^T F_k$. Se $g_k = 0$, parar.

Passo 2: Obter $W_k \in \mathbb{R}^n$ tal que

$$\| J_k^T J_k W_k + J_k^T F_k \| \leq \eta_k \| J_k^T F_k \|. \qquad (3.2.1)$$

Obs: Isto é feito usando-se o algoritmo de gradientes conjugados aplicado ao problema :

$$\min_W \frac{1}{2} \| J_k W + F_k \|_2^2,$$

com precondicionamento de J_k.

Passo 3: Obter $V_k \in \mathbb{R}^n$, solução do seguinte problema bidimensional:

$$\text{Minimizar } \| J_k V + F_k \|$$
$$\text{suj. a } V = \lambda_1 g_k + \lambda_2 W_k, \quad \| V \| \leq \| W_k \|.$$

Passo 4: Seja $d_k^1 = -g_k$. Se V_k cumpre as duas condições seguintes:

$$< V_k , g_k > \leq -\theta_1 \| V_k \| \| g_k \|, \tag{3.2.2}$$

$$\underline{M} \| g_k \| \leq \| V_k \| \leq \bar{M} \| g_k \|, \tag{3.2.3}$$

então $d_k^2 = V_k$. Caso contrário $d_k^2 = d_k^1$.

Passo 5: Seja $t = \| d_k^2 \|$. Executar os passos de (5.a) até (5.d):

(5.a)_ Resolver o problema

$$\text{Minimizar } \| J_k d + F_k \| \tag{3.2.4}$$
$$\text{suj. a } d = \lambda_1 d_k^1 + \lambda_2 d_k^2 , \| d \| \leq t.$$

(5.b)_ Se $\frac{1}{2} \| F(x_k + d) \|^2 \leq \frac{1}{2} \| F(x_k) \|^2 + \theta_2 < g_k , d >$, (3.2.5)

passar ao passo (5.d).

(5.c)_ Obter $\bar{t}$ tal que

$$\theta_3 t \leq \bar{t} \leq (1 - \theta_3) t.$$

Substituir t por $\bar{t}$ e voltar ao passo (5.a).

(5.d)_ $d_k = d$, $x_{k+1} = x_k + d_k$.

Provaremos abaixo que este algoritmo está bem definido e para isto mostraremos que a condição (3.2.5) é satisfeita para t bastante pequeno.

Consideremos a seguinte função:

$$\varphi(d) = \frac{1}{2} \frac{\| F(x_k + d) \|^2 - \| F(x_k) \|^2}{\| d \|}$$

Temos que, pelo teorema do valor médio,

$$\varphi(d) = \frac{< g(x_k + \xi d) , d >}{\| d \|} , \quad 0 \leq \xi \leq 1.$$

Se d(t) é solução do problema (3.2.4), problema de duas variáveis, restrito a uma bola bidimensional de raio t, então

$$\lim_{t \to 0} \frac{d(t)}{\| d(t) \|}$$

é a direção de máxima descida da função restrita ao plano gerado por d_k^1 e d_k^2. Mas a direção de máxima descida da função (irrestrita) pertencente a este plano ($d_k^1 = - g_k$). Portanto,

$$\lim_{t \to 0} \frac{d(t)}{\| d(t) \|} = \frac{- g_k}{\| g_k \|} .$$

Por outro lado, quando t tende a 0, $\varphi(d)$ tende a $-\| g_k \|$ e $< g_k , \frac{d}{\| d \|} >$ também tende a $-\| g_k \|$. Portanto,

$$\text{Lim}\ \frac{1}{2}\ \frac{\| F(x_k + d) \|^2 - \| F(x_k) \|^2}{< g_k , d >} = 1,$$

isto implica que

$$\frac{1}{2}\ \frac{\| F(x_k + d) \|^2 - \| F(x_k) \|^2}{< g_k , d >} \geq \theta_2,$$

se t é bastante pequeno. Mas $< g_k , d > < 0$; assim a desigualdade (3.2.5) ocorre para t pequeno.

3.3. Resultados de Convergência Global

Nesta seção chamaremos de f(x) a $\frac{1}{2} \| F(x) \|_2^2$. De todas as maneiras, é fácil ver que os resultados se aplicam a qualquer $f \in C^1(D)$, com $\nabla f(x) = g(x)$.

Admitindo as hipóteses consideradas na seção anterior e sendo $\mu > 1$, observando o algoritmo (3.2), verificaremos que este é um caso particular do apresentado abaixo.

Obs: Definimos $C(v, w) = \{ x \in \mathbb{R}^n / x = \gamma_1 v + \gamma_2 w;\ \gamma_1, \gamma_2 \geq 0 \}$, cone convexo determinado por v e w.

Algoritmo 3.3

Dado $x_0 \in D$, um ponto inicial, arbitrário,

consideremos a sequência definida da seguinte maneira:

se $g_k = g(x_k) = 0$, parar. Caso contrário, calcular

$$x_{k+1} = x_k + d_k$$

onde

$$d_k \in C(d_k^1, d_k^2), \quad \| d_k \| \leq \| d_k^2 \|, \tag{3.3.1}$$

e satisfaz as seguintes condiçoẽs:

$$< d_k^i , g_k > \leq - \theta_1 \| d_k^i \| \| g_k \|, \quad i = 1,2, \tag{3.3.2}$$

$$\underline{M} \| g_k \| \leq \| d_k^i \| \leq \bar{M} \| g_k \|, \quad i = 1,2, \tag{3.3.3}$$

$$f(x_k + d_k) \leq f(x_k) + \theta_2 < g_k , d_k >. \tag{3.3.4}$$

Finalmente uma das seguintes possibilidades é verdadeira:

$$d_k = d_k^2 \tag{3.3.5}$$

ou

Existe $\bar{d}_k \in C(d_k^1, d_k^2)$, $\| \bar{d}_k \| \leq \mu \| d_k \|$

e $$f(x_k + \bar{d}_k) > f(x_k) + \theta_2 < g_k , d_k > \tag{3.3.6}$$

TEOREMA 3.3.1: Se $x_* \in D$ é um ponto limite da sequência gerada pelo algoritmo 3.3, então $g(x_*) = 0$.

Demonstração: Se x_* é um ponto limite da sequência (x_k), então existe uma subsequência (x_k), $k \in K_1 \subset \mathbb{N}$ tal que:

$$\lim_{k \in K_1} x_k = x_*$$

Definimos $B = \{ x_k, k \in K_1 \}$. B é um subconjunto limitado de D. Portanto, como $f \in C^1(D)$, temos

$$\| g(x) \| \leq C_1 \text{ para todo } x \in B.$$

Suponhamos que $g(x_*) \neq 0$. Então, existe K_2 um subconjunto infinito de K_1, tal que:

$$\| g(x_k) \| \geq C_2 > 0, \text{ para todo, } k \in K_2.$$

Portanto

$$\underline{M}\, C_2 \leq \| d_k^i \| \leq \bar{M}\, C_1, \text{ para todo } k \in K_2, \; i = 1, 2.$$

Logo, existe $K_3 \subset K_2$ tal que

$$\lim_{k \in K_3} d_k^i = d^i, \; i = 1, 2$$

e

$$\underline{M}\, C_2 \leq \| d^i \| \leq \bar{M}\, C_1, \; i = 1, 2.$$

Tomando o limite em ambos os lados de (3.3.2) e (3.3.3) para $k \in$

K_3, obtemos:

$$< d^1 , g(x_*) > \leq - \theta_1 \| d^1 \| \| g(x_*) \|$$

$$\underline{M} \| g(x_*) \| \leq \| d^1 \| \leq \bar{M} \| g(x_*) \|, \quad 1 = 1, 2.$$

Consideremos duas possibilidades (a) e (b)

a) Existe $\alpha > 0$ tal que $\| d_k \| \geq \alpha \| g_k \|$ para todo $k \in K_3$

b) O oposto de (a)

(a) Neste caso

$$\alpha C_2 \leq \alpha \| g(x_k)\| \leq \| d_k \| \leq \| d_k^2 \| \leq \bar{M} \| g(x_k) \| \leq \bar{M} C_1$$

para $k \in K_3$.

Portanto, existe $K_4 \subset K_3$ tal que

$$\lim_{k \in K_4} d_k = d \neq 0.$$

Por (3.3.1), temos que d_k está no cone positivo determinado por d_k^1 e d_k^2. Um rápido cálculo mostra que a desigualdade (3.3.2) ocorre para todos os membros do cone. Então:

$$< d_k , g(x_k) > \leq -\theta_1 \| d_k \| \| g_k \|.$$

Tomando o limite em ambos os lados da desigualdade para $k \in K_4$, obtemos

$$< d , g(x_*) > \leq -\theta_1 \| d \| \| g(x_*) \| = - \gamma < 0.$$

Tomando também o limite em ambos os lados de (3.3.4), obtemos

$$f(x_* + d) = \lim_{k \in K_4} f(x_k + d_k) \leq \lim_{k \in K_4} f(x_k) + \theta_2 < g(x_k) , d_k >$$

$$= f(x_*) + \theta_2 < g(x_*) , d >$$

$$\leq f(x_*) - \theta_1 \theta_2 \| d \| \| g(x_*) \|$$

$$= f(x_*) - \theta_2 \gamma$$

Portanto, existe $k_0 \in \mathbb{N}$ tal que, se $k \geq k_0$, $k \in K_4$

$$f(x_{k+1}) = f(x_k + d_k) \leq f(x_*) - \theta_2 \gamma/2,$$

o que é uma contradição.

(b) Neste caso, $d_k \neq d_k^2$ e $\lim_{k \in K_5} d_k = 0$ para todo $k \in K_5$, um subconjunto infinito de K_3. Portanto,

$$\lim_{k \in K_5} \bar{d}_k = 0$$

e

$$f(x_k + \bar{d}_k) > f(x_k) + \theta_2 < g_k , \bar{d}_k >.$$

Assim, para $k \in K_5$,

$$< g(x_k + \xi_k \bar{d}_k) , \bar{d}_k > > \theta_2 < g_k , \bar{d}_k >, \ 0 \leq \xi_k \leq 1,$$

de forma que

$$< g(x_k + \xi_k \bar{d}_k), \frac{\bar{d}_k}{\| \bar{d}_k \|} > > \theta_2 < g_k , \frac{\bar{d}_k}{\| \bar{d}_k \|} >. \quad (3.3.7)$$

Agora seja K_6 um subconjunto infinito de K_5, tal que

$$\lim_{k \in K_6} \frac{\bar{d}_k}{\| \bar{d}_k \|} = V.$$

Tomando os limites em ambos os lados de (3.3.7), para $k \in K_6$, temos

$$< g(x_*) , V > \geq \theta_2 < g(x_*) , V >.$$

Mas $\bar{d}_k$ pertence a $C(d_k^1 , d_k^2)$, logo

$$< \frac{\bar{d}_k}{\| \bar{d}_k \|} , g(x_k) > \leq \theta_1 \| g(x_k) \|.$$

Portanto, $< V , g(x_*) > < 0$. Isto implica que $\theta_2 > 1$, o que também é uma contradição. ■

COROLÁRIO 3.3.2: Seja $\varepsilon > 0$. Se $\{ x: f(x) \leq f(x_0) \}$ é compacto, então existe $k \in \mathbb{N}$ tal que $\| g(x_k) \| \leq \varepsilon$.

LEMA 3.3.3: Seja x_* um mínimo local estrito de f em D, $\varepsilon > 0$. Então existe uma vizinhança V de x_* tal que $x_k \in B(x_*, \varepsilon)$ para todo $k \geq k_0$, desde que $x_{k_0} \in V$.

TEOREMA 3.3.4: Se x_* é um mínimo local estrito de f em D, então existe $\varepsilon > 0$ tal que Lim $x_k = x_*$, para $x_0 \in B(x_*, \varepsilon)$.

3.4 Estratégia de Busca no Plano Bidimensional

Consideremos o caso usual no qual $d_k^1 \neq d_k^2$. Seja $\{ e_1, e_2 \}$ uma base ortonormal do subespaço gerado por $\{ d_k^1, d_k^2 \}$, e seja $E = (e_1, e_2)$.

Portanto, o problema

$$\text{Minimizar } \| J_k W + F_k \|^2 \qquad (3.5.1)$$

$$\text{suj. a } W = E \alpha \, , \; \| W \| \leq t$$

pode ser escrito na forma

$$\text{Minimizar } \| J_k E \alpha + F_k \|$$

$$\text{suj. a } \| \alpha \| \leq t,$$

que podemos resolver como um problema bidimensional.

Chamando de W(t) a solução de (3.5.1), então, se W(t) não satisfizer a condição (5.b) do algoritmo 3.2, o valor de t é diminuído e volta-se a resolver (3.5.1). A atualização de t é feita por interpolação cúbica: Definindo

$$\sigma(t) = \| F(x_k + W(t) \|^2$$

calculamos $\sigma(0)$, $\sigma'(0)$, $\sigma(t)$, $\sigma'(t)$ e interpolamos, esperando que o valor de t esteja no intervalo [0.1t, 0.9t]. Senão, substituímos t por (0.1t) ou (0.9t).

CAPÍTULO IV

MODIFICAÇÕES DA ESTRATÉGIA BIDIMENSIONAL

4.1. Introdução

No capítulo anterior mostramos um algoritmo que tenta manter as boas características dos métodos Gauss - Newton e Levenberg - Marquardt e ter um desempenho melhor nos casos em que estes métodos não tenham um bom comportamento.

Neste capítulo, tentamos melhorar o algoritmo 3.2 em dois sentidos. Em primeiro lugar, melhorar o desempenho em problemas mal escalados, através da substituição de ∇f por $D\nabla f$, um gradiente escalado. Em segundo lugar, simplificar o processo de busca, introduzindo um processo de busca parabólica, em vez do processo de região de confiança. Provamos que a convergência local e global continua válida para esta modificação.

4.2 Problemas Mal Escalados

Os problemas mal escalados surgem frequentemente em situações práticas, e é nosso objetivo melhorar o algoritmo 3.2 neste sentido, sem que a modificação que faremos possa ser prejudicial em outros problemas que não tenham esta característica. Isto é, desejamos trabalhar com um gradiente

convenientemente escalado, mas que o comportamento do algoritmo seja indiferente nos casos gerais.

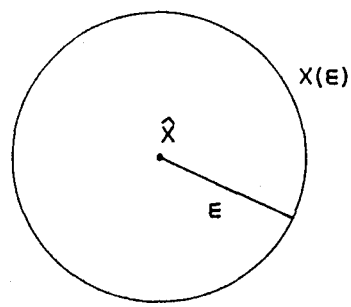
Consideremos o problema:

$$\text{Min} \quad f(x)$$

$$\text{suj. a} \quad \| x - \hat{x} \|^2 = \varepsilon^2,$$

onde $f: \mathbb{R}^n \longrightarrow \mathbb{R}$, $x \in \mathbb{R}^n$, $g(x) = \nabla f(x)$ e $g(\hat{x}) \neq 0$. Sabemos a direção de máxima descida é $-g(x)$. Vejamos o porquê.

Se a solução está em $x(\varepsilon)$,



temos por Lagrange que:

$$g(x(\varepsilon)) + \lambda \begin{bmatrix} x_1 - \hat{x}_1 \\ \vdots \\ x_n - \hat{x}_n \end{bmatrix} = 0, \quad \lambda \in \mathbb{R}$$

$$\Longrightarrow \begin{bmatrix} x_1 - \hat{x}_1 \\ \vdots \\ x_n - \hat{x}_n \end{bmatrix} = \bar{\lambda} \begin{bmatrix} g_1(x(\varepsilon)) \\ \vdots \\ g_n(x(\varepsilon)) \end{bmatrix}, \quad \text{onde } \bar{\lambda} = -1/\lambda$$

Portanto, para um ε muito pequeno, $x - \hat{x}$ tem a direção de $-g(\hat{x})$.

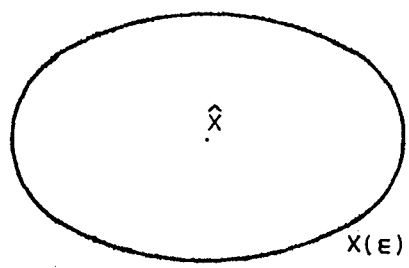
Agora consideremos que queremos minimizar a mesma

função, mas sujeita a outra restrição, ou seja:

$$\text{Min } f(x)$$

$$\text{suj. a } \frac{(x_1 - \hat{x}_1)^2}{d_1^2} + \ldots + \frac{(x_n - \hat{x}_n)^2}{d_n^2} = \varepsilon^2.$$

De modo equivalente ao problema anterior, temos, por Lagrange, que:



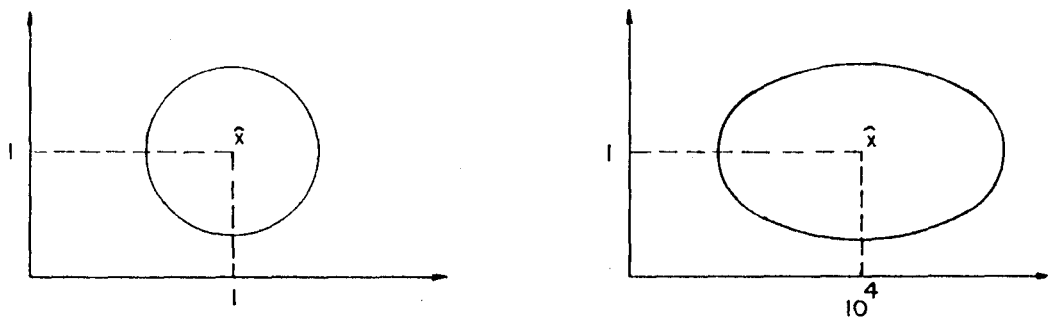
$$g(x(\varepsilon)) + \lambda \begin{bmatrix} \frac{x_1 - \hat{x}_1}{d_1^2} \\ \vdots \\ \frac{x_n - \hat{x}_n}{d_n^2} \end{bmatrix} = 0$$

$$\Longrightarrow \begin{bmatrix} x_1 - \hat{x}_1 \\ \vdots \\ x - \hat{x} \end{bmatrix} = -\bar{\lambda} \begin{bmatrix} d_1^2 & & \\ & \ddots & \\ & & d_n^2 \end{bmatrix} g(x(\varepsilon)).$$

Portanto, a direção $-Dg(\hat{x})$ deve ser considerada a direção de máxima descida relativa à norma definida por:

$$\| x - \hat{x} \|_D^2 = \frac{(x_1 - \hat{x}_1)^2}{d_1^2} + \ldots + \frac{(x_n - \hat{x}_n)^2}{d_n^2}$$

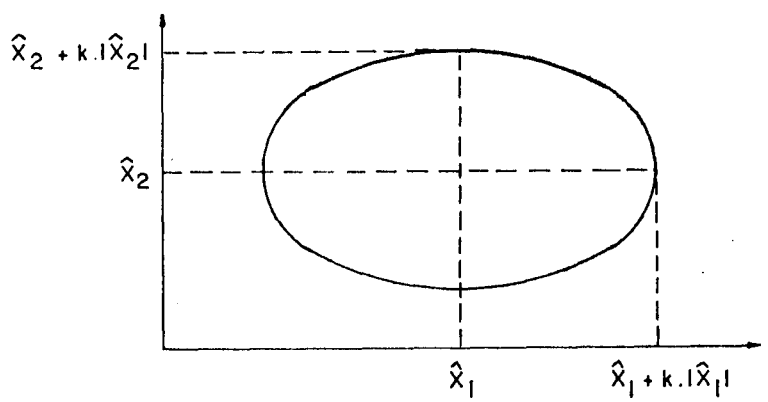
Como podemos observar, as direções de máxima descida que tomamos dependem da norma definida no problema. O que nos leva a considerar normas, e portanto direções de máxima descida, diferentes da usual é o " scaling " de $\hat{x}$. Se algumas das coordenadas de $\hat{x}$ são muito maiores que outras, é provável que estejam medidas em unidades diferentes. Como consequência disto, é razoável que as " grandes coordenadas " variem mais que as " pequenas coordenadas ". Ilustrando temos:



Norma Usual

Norma modificada por scaling

Logo, de um modo geral, parece razoável que a variação em cada semi-eixo seja proporcional ao tamanho de $\hat{x}_1$. Isto nos leva a

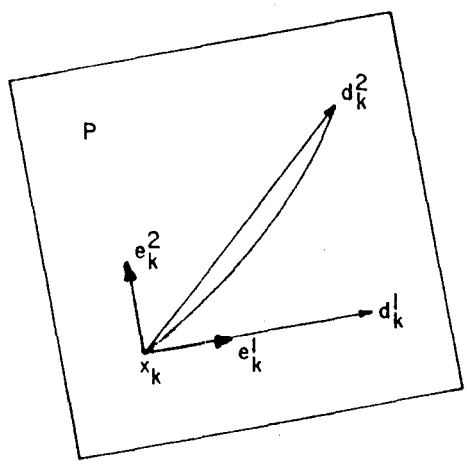


donde podemos considerar a norma $\| \cdot \|_D$ com $d_1 = |\hat{x}_1|$.

4.3. Busca na Parábola

A modificação que faremos no algoritmo 3.2, deixando de trabalhar com a estratégia de região de confiança para trabalharmos com uma busca na parábola, foi feita no sentido de diminuir o custo computacional do algoritmo original e simplificar o cálculo do novo ponto.

Suponhamos que estejamos no passo 5 do algoritmo 3.2: consideremos o plano gerado por d_k^1 e d_k^2 ; pelo processo de Gram - Schmidt obtemos $\{ e_k^1 , e_k^2 \}$



$$e_k^1 = \frac{d_k^1}{\| d_k^1 \|}$$

$$\tilde{e}_k^2 = d_k^2 - < d_k^2 , e_k^1 > e_k^1$$

$$e_k^2 = \frac{\tilde{e}_k^2}{\| \tilde{e}_k^2 \|}$$

Portanto se $x \in P$, x é escrito da seguinte forma

$$x = x_k + y_1 e_k^1 + y_2 e_k^2 ;$$

chamando $E = (e_k^1 , e_k^2)$ isto implica que $x = x_k + E\,y$, logo $F(x) = F(x_k + Ey)$.

Mas o que queremos é uma parábola que seja tangente a d_k^1 e passe por d_k^2 . Portanto, ela deve ter a forma

$$y_2 = a\,y_1^2 \Longrightarrow \qquad a = \frac{y_2}{y_1^2} .$$

Mas no passo 4 obtemos uma direção que é de descida, logo a nossa primeira direção a tomar é a própria d_k^2 . Assim temos

$\Longrightarrow x_k + E\,y = x_k + d_k^2$

$\Longrightarrow E\,y = d_k^2$

$\Longrightarrow y_1\,e_k^1 + y_2\,e_k^2 = d_k^2$

$\Longrightarrow y_1\,e_k^1 + y_2\,e_k^2 = < d_k^2 , e_k^1 > e_k^1 + < d_k^2 , e_k^2 > e_k^2$

$\Longrightarrow y_1 = < d_k^2 , e_k^1 > \quad e \quad y_2 = < d_k^2 , e_k^2 > .$

Esses são, portanto, os valores de y_1 e y_2 que correspondem a d_k^2.

Como podemos observar, os parâmetros da parábola são facilmente calculados.

4.4. Algoritmo Modificado

Vamos admitir as mesmas hipóteses consideradas para o algoritmo 3.2 na seção 3.2.

Algoritmo 4.1

Passo 1: Computar J_k e g_k. Se $g_k = 0$, parar

Passo 2: Obter $w_k \in \mathbb{R}^n$ tal que

$$\| J_k^T J_k w_k + g_k \| \leq \eta_k \| g_k \|.$$

Obs: obter $w_k = \arg \min \| J_k w + F_k \|$ utilizando o algoritmo de gradientes conjugados com precondicionamento de J_k.

Passo 3: Obter V_k como solução de

$$\text{Min} \| J_k V + F_k \|_2^2$$

$$\text{Suj. a } V = \lambda_1 g_k + \lambda_2 w_k, \ \| V \| \leq \| w_k \|.$$

Passo 4: $d_k^1 = - D g_k$

Se V_k satisfizer as condições abaixo,

$$< V_k , g_k > \leq - \theta_1 \| V_k \| \| g_k \| \quad \text{e}$$

$$\underline{M} \| g_k \| \leq \| V_k \| \leq \bar{M} \| g_k \|,$$

então $d_k^2 = V_k$. Caso contrário $d_k^2 = d_k^1$.

Passo 5:

(5.a): Considerar o plano gerado por d_k^1 e d_k^2. Calcular pelo processo de Gram - Schmidt $\{ e_1 , e_2 \}$, para depois calcular os parâmetros

$$y_1 = < d_k^2 , e_k^1 >$$

$$y_2 = < d_k^2 , e_k^2 >$$

da parábola $y_2 = a\ y_1^2$.

Como x pertence à parábola, tem a forma

$$x = x_k + e_k^1\ y_1 + e_k^2\ a\ y_1^2$$

logo,

$$\| F(x) \|^2 = \| F(x_k + e_k^1\ y_1 + e_k^2\ a\ y_1^2) \|^2 = \varphi(y_1).$$

Obs: chamaremos $t = y_1$

(5.b): Calcular $\varphi(t)$

se $\frac{1}{2}\ \varphi(t) \leq \frac{1}{2}\ \varphi(0) + \theta_2 < g_k , d >$ ir para (5.d)

Obs: $d = e_k^1\ t + e_k^2\ a\ t^2$

(5.c): $t \longleftarrow t/2$. Voltar ao passo (5.b)

(5.d): $d_k = d$ e $x_{k+1} = x_k + d$

Obs: este algoritmo segue quase os mesmos passos do algoritmo 3.2, mas com algumas modificações nos passos 4 e 5.

4.5 Consistência Teórica

Nesta seção mostramos que, apesar das mudanças feitas no algoritmo 3.2, no sentido de melhorar o desempenho em problemas mal escalados e introduzir um processo de busca mais simples, como a busca parábolica, não foram afetados os resultados de convergência local e global.

TEOREMA 4.5.1: Dado $x^* \in D$ (conjunto aberto contido no $\mathbb{R}^n$), um ponto limite da sequência gerada pelo algoritmo 4.1, então $g(x^*) = 0$.

No capítulo anterior provamos que o algoritmo 3.2 é um caso particular do algoritmo 3.3; na demonstração do teorema 3.3.1 podemos observar que independentemente do modo que calculamos a direção, o que é comum nestes algoritmos é a região em que trabalhamos no cálculo da direção, que é o cone convexo formado por d_k^1 e d_k^2. Do mesmo modo podemos dizer que o algoritmo 4.1 é um caso especial do algoritmo 3.3.

CAPÍTULO V

EXPERIÊNCIAS NUMÉRICAS

5.1. Introdução

As experiências foram feitas com o algoritmo original e o algoritmo modificado. Mas houve também uma variação no algoritmo modificado na atualização do parâmetro "t", já que na primeira versão é usada a bissecção e na segunda é usada a interpolação cúbica.

Trabalhamos com dois tipos de convergência:

1 - $\| F(x) \| \leq$ sup (sup - mínimo conhecido ou um valor razoavelmente pequeno).

2 - $\| g(x) \| \leq$ eps (eps - valor muito pequeno).

As experiências foram feitas em um equipamento VAX/VMS do CCUEC/ UNICAMP usando a linguagem FORTRAN.

5.2. Experimentos

As funções testes são as seguintes:

Obs: f é o valor exato do resíduo.

(1) Função de Rosenbrock

$m = 2$, $n = 2$

$F_1(x) = 10(x_2 - x_1^2)$

$F_2(x) = 1 - x_1$

$x_0 = (-5, 1)$

$f = 0$

(2) Função Mal Escalada de Powell

$m = 2$, $n = 2$

$F_1(x) = 10^4 x_1 x_2 - 1$

$F_2(x) = \exp[-x_1] + \exp[-x_2] - 1.0001$

$x_0 = (0, 1)$

$f = 0$

(3) Função de Beale

$m = 3$, $n = 2$

$F_i(x) = Y_i - x_1(1 - x_2^i)$

Onde $Y_1 = 1.5$, $Y_2 = 2.25$, $Y_3 = 2.625$

$x_0 = (1, 1)$

$f = 0$ em $(3, 0.5)$

(4) Função de Jennrich e Sampson

$m = 10$, $n = 2$

$F_i(x) = 2 + 2i - (\exp[ix_1] + \exp[ix_2])$

$x_0 = (0.3, 0.4)$

$f = 124.36$

(5) Função Helical Valley

$m = 3$, $n = 3$

$F_1(x) = 10[x_3 - 10\,\theta(x_1, x_2)]$

$F_2(x) = 10[(x_1^2 + x_2^2)^{1/2} - 1]$

$F_3(x) = x_3$

Onde

$$\theta(x_1, x_2) = \begin{cases} \frac{1}{2\pi} \text{ arctg } \left[\frac{x_2}{x_1}\right] & \text{se } x_1 > 0 \\ \frac{1}{2\pi} \text{ arctg } \left[\frac{x_2}{x_1}\right] + 0.5 & \text{se } x_1 < 0 \end{cases}$$

$x_0 = (-1, 0, 0)$

$f = 0$ em $(1, 0, 0)$

(6) Função de Bard

$m = 15$, $n = 3$

$$F_i(x) = Y_i - \left[x_1 + \frac{u_i}{v_i x_2 + w_i x_3} \right]$$

Onde $u_i = i$, $v_i = 16 - i$, $w_i = \min(u_i, v_i)$ e

i	Y	i	Y	i	Y
1	0.14	6	0.32	11	0.73
2	0.18	7	0.35	12	0.96
3	0.22	8	0.39	13	1.34
4	0.25	9	0.37	14	2.10
5	0.29	10	0.58	15	4.39

$x_0 = (1, 1, 1)$

$f = 8.215e\text{-}3$

(7) Função Gaussiana

$m = 15$, $n = 3$

$$F_i(x) = x_1 \exp\left[\frac{-x_2(t_i - x_3)^2}{2}\right] - Y_i$$

Onde $t_i = (8 - i)/2$ e

i	Y
1, 15	0.0009
2, 14	0.0044
3, 13	0.0175
4, 12	0.0540
5, 11	0.1295
6, 10	0.2420
7, 9	0.3521
8	0.3989

$x_0 = (0.4,\ 1,\ 0)$

$f = 1.128e\text{-}8$

(8) Função de Desenvolvimento e Pesquisa de Gulf

$m = 6$, $n = 3$

$$F_i(x) = \exp\left[-\frac{\left|Y_i - m_i\right|^{x_3}}{x_1}\right] - t_i$$

Onde $t_i = i/100$ e

$Y_i = 25 + (-50 \ln(t_i))^{2/3}$

$x_0 = (5,\ 2.5,\ 0.15)$

$f = 0$

(9) Função de Box Tridimensional

$m = 9$, $n = 3$

$F_i(x) = \exp[-t_i x_1] - \exp[-t_i x_2] - x_3(\exp[-t_i] - \exp[-10t_i])$

Onde $t_i = (0.1)i$

$x_0 = (0,\ 10,\ 20)$

$f = 0$ em $(1,\ 10,\ 1)$

(10) Função Singular de Powell

$m = 4\ ,\ n = 4$

$F_1(x) = x_1 + 10x_2$

$F_2(x) = 5^{1/2}(x_3 - x_4)$

$F_3(x) = (x_2 - 2x_3)^2$

$F_4(x) = 10^{1/2}(x_1 - x_4)^2$

$x_0 = (3,\ -1,\ 0,\ 1)$

$f = 0$

(11) Função de Wood

$m = 6\ ,\ n = 4$

$F_1(x) = 10(x_4 - x_1^2)$

$F_2(x) = 1 - x_1$

$F_3(x) = (90)^{1/2}(x_4 - x_3^2)$

$F_4(x) = 1 - x_3$

$F_5(x) = (10)^{1/2}(x_2 + x_4 - 2)$

$F_6(x) = (10)^{-1/2}(x_2 - x_4)$

$x_0 = (0.25,\ 0.39,\ 0.415,\ 0.39)$

$f = 0$ em $(1,\ 1,\ 1,\ 1)$

(12) Função de Kowalik e Osborne

$m = 11$, $n = 4$

$$F_i = Y_i - \frac{x_1(u_i^2 + u_i x_2)}{(u_i^2 + u_i x_3 + x_4)}$$

Onde

i	Y	u	i	Y	u
1	0.1957	4.0000	7	0.0456	0.1250
2	0.1947	2.0000	8	0.0342	0.1000
3	0.1735	1.0000	9	0.0323	0.0833
4	0.1600	0.5000	10	0.0235	0.0714
5	0.0844	0.2500	11	0.0246	0.0625
6	0.0627	0.1670			

$x_0 = (0.25,\ 0.39,\ 0.415,\ 0.39)$

$f = 3.075e\text{-}4$

(13) Função 1 de Osborne

$m = 33$, $n = 5$

$$F_i(x) = Y_i - \left(x_1 + x_2 \exp\left[-t_i x_4\right] + x_3 \exp\left[-t_i x_5\right]\right)$$

Onde $t_i = 10(i - 1)$ e

i	Y	i	Y	i	Y
1	0.844	12	0.718	23	0.478
2	0.908	13	0.685	24	0.467
3	0.932	14	0.658	25	0.457
4	0.936	15	0.628	26	0.448
5	0.925	16	0.603	27	0.438
6	0.908	17	0.580	28	0.431
7	0.881	18	0.558	29	0.424
8	0.850	19	0.538	30	0.420
9	0.818	20	0.522	31	0.414
10	0.784	21	0.506	32	0.411
11	0.751	22	0.490	33	0.406

$x_0 = (0.5,\ 1.5,\ -1,\ 0.01,\ 0.02)$

$f = 5.46489e\text{-}5$

(14) Função Singular de Powell Estendida

$m = n\ ,\ n = 16$

$F_{4i-3}(x) = x_{4i-3} + 10\ x_{4i-2}$

$F_{4i-2}(x) = 5^{1/2}(x_{4i-1} - x_{4i})$

$F_{4i-1}(x) = (x_{4i-2} - 2\ x_{4i-1})^2$

$F_{4i}(x) = 10^{1/2}(x_{4i-3} - x_{4i})^2$

$x_0 = (\xi)$

Onde $\xi_{4i-3} = 3,\ \xi_{4i-2} = -1,\ \xi_{4i-1} = 0,\ \xi_{4i} = 1$

$f = 0$

(15) Função Variável Dimensional

$m = n + 2\ ,\ n = 4$

$F_i(x) = x_i - 1 \quad i = 1, \ldots, n$

$$F_{n+1}(x) = \sum_{j=1}^{n} j(x_j - 1)$$

$$F_{n+2}(x) = \left[\sum_{j=1}^{n} j(x_j - 1) \right]^2$$

$x_0 = (\xi)$ onde $\xi_j = 1 - (j/n)$

$f = 0$ em $(1, \ldots, 1)$

(16) Função Trigonométrica

$m = n$, $n = 6$

$$F_i(x) = n - \sum_{j=1}^{n} \cos x_j + i(1 - \cos x_i) - \operatorname{sen} x_i$$

$x_0 = (1/n, \ldots, 1/n)$

$f = 0$

(17) Função Valor Fronteira Discreta

$m = n$, $n = 5$

$$F_i(x) = 2x_i - x_{i-1} - x_{i+1} + h^2(x_i + t_i + 1)^3/2$$

Onde $h = 1/(n + 1)$, $t_i = ih$, e $x_0 = x_{n+1} = 0$

$x_0 = (1, \ldots, 1)$

$f = 0$

(18) Função Banda de Broyden

$m = n$, $n = 6$

$$F_i(x) = x_i(2 + 5x_i^2) + 1 - \sum_{j \in J_i} x_j(1 + x_j)$$

Onde $J_i = \{j: j \neq i, \max(1, i - m_i) \leq j \leq \min(n, i + m_u)\}$

$x_0 = (-1, \ldots, -1)$

$f = 0$

(19) Função de Watson

$m = 31$, $n = 9$

$$F_i(x) = \sum_{j=2}^{n} (j - 1)\, x_j\, t_i^{j-2} - \left[\sum_{j=1}^{n} x_j t_i^{j-1} \right]^2 - 1$$

Onde $t_i = i/29$, $1 \leq i \leq 29$

$F_{30}(x) = x_1$

$F_{31}(x) = x_2 - x_1^2 - 1$

$x_0 = (1.e\text{-}6, \ldots, 1.e\text{-}6)$

$f = 1.39976e\text{-}6$ p/ $n = 9$

(20) Função Equação Integral Discreta

$m = n$, $n = 10$

$$F_i(x) = x_i + h \left[(1 - t_i) \sum_{j=1}^{i} t_j (x_j + t_j + 1)^3 + t_i \sum_{j=i+1}^{n} (1 - t_j)(x_j + t_j + 1)^3 \right] / 2$$

Onde $h = 1/(n + 1)$, $t_i = ih$, e $x_0 = x_{n+1} = 0$

$x_0 = (1.e\text{-}6, \ldots, 1.e\text{-}6)$

$f = 0$

Os resultados são apresentados na tabelas seguintes, onde:

AO = algoritmo Original(3.2).

AM = algoritmo modificado(4.1).

AMC = algoritmo modificado atualizando o parâmetro t usando interpolação cubíca.

NI = número de iterações

NA = número de avaliações

RE = resíduo

NF = número da função teste

TABELA 1

	AO			AM		
NF	NI	NA	RE	NI	NA	RE
1	6	7	0.0	6	7	0.0
2	7	9	0.48E-05	7	9	0.17E-04
3	10	12	0.0	8	9	0.0
4	8	24	124.3	21	145	124.9
5	21	34	0.0	11	12	0.06
6	6	6	0.82E-02	6	6	0.82E-02
7	3	3	0.11E-07	3	3	0.11E-07
8	24	24	0.86E-02	2	2	0.91E-02
9	7	7	0.0	7	7	0.0
10	15	15	0.35E-13	15	15	0.22E-14
11	9	12	0.0	7	8	0.0
12	8	12	0.30E-03	5	7	0.30E-03
13	10	15	0.54E-04	3	8	0.12E-03
14	2	2	40.24	2	2	40.25
15	9	9	0.0	7	29	0.0
16	11	18	0.30E-13	43	49	0.71E-14
17	5	5	0.38E-18	5	5	0.50E-17
18	7	7	0.84E-13	7	7	0.84E-13
19	5	5	0.13E-05	12	12	0.14E-05
20	9	10	0.55E-16	9	10	0.55E-16

TABELA 2

	AMC			AM		
NF	NI	NA	RE	NI	NA	RE
1	8	12	0.0	6	7	0.0
2	7	9	0.17E-04	7	9	0.17E-04
3	13	15	0.0	8	9	0.0
4	21	70	124.9	21	145	124.9
5	11	12	0.0	11	12	0.0
6	8	10	0.82E-02	6	6	0.82E-02
7	3	3	0.11E-07	3	3	0.11E-07
8	2	2	0.91E-02	2	2	0.91E-02
9	7	7	0.0	7	7	0.0
10	15	15	0.22E-14	15	15	0.22E-14
11	7	8	0.0	7	8	0.0
12	15	30	0.30E-03	5	7	0.30E-03
13	3	6	0.13E-04	3	8	0.12E-03
14	2	2	40.25	2	2	40.25
15	7	29	0.0	7	29	0.0
16	4	129	0.19E-02	43	49	0.71E-14
17	5	5	0.50E-18	5	5	0.50E-17
18	7	7	0.84E-13	7	7	0.84E-13
19	12	12	0.14E-05	12	12	0.14E-05
20	9	10	0.55E-16	9	10	0.55E-16

CONCLUSÃO

1) Na maioria das funções os algoritmos tiveram desempenhos iguais , o número de avaliações foi igual ou quase igual ao número de iterações. Apesar dos dois algoritmos terem os mesmos comportamentos podemos pensar que nestes casos o algoritmo modificado é levemente superior ao original devido às suas iterações serem mais simples.

2) Em algumas funções o algoritmo modificado teve desempenho melhor do que o do original, como na função de Wood, onde tanto o número de iterações como o número de avaliações são menores quando usamos o algoritmo modificado em vez do algoritmo original, como também na função Helical Valley em que o desempenho do algoritmo modificado é superior ao do algoritmo original.

3) Houve casos em que o algoritmo original foi melhor do que o modificado, como por exemplo, no caso da função de Watson.

4) Também temos casos em que nenhum dos algoritmos funcionou, como no caso da função singular de Powell Estendida.

5) Com relação ao algoritmo modificado com atualização do parâmetro t por bissecção ou interpolação cúbica seus desempenhos foram quase iguais.

Como podemos observar não fizemos testes para problemas de grande porte, portanto para pesquisas futuras restaria analisar mais os dois algoritmos para podermos empregá-los neste tipo de problemas.

APÊNDICE A

LEMA 1: Seja $F: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ contínua, diferenciavel em um conjunto convexo aberto $D \subset \mathbb{R}^n$, $x \in D$, e seja J Lipschitz contínua em x na vizinhança de D, usando uma norma vetorial e a norma operador matricial induzida e a constante γ. Então, para todo $x + P \in D$,

$$\| F(x + P) - F(x) - J(x)P \| \leq \frac{\gamma}{2} \| P \|^2.$$

LEMA 2: Seja $F: \mathbb{R}^n \longrightarrow \mathbb{R}$, contínua, duas vezes diferenciáveis, seja $H \in \mathbb{R}^{n \times n}$, simétrica definida positiva e seja $\| \cdot \| = \| \cdot \|_2$. Então o problema

$$\begin{aligned} &\text{Minimizar } F(x) + \nabla F(x)^T S + \frac{1}{2} S^T H S \\ &\text{Suj. a} \quad \| S \| \leq \delta \end{aligned}$$

é resolvido por

$$S(\mu) \stackrel{\Delta}{=} - (H + \mu I)^{-1} \nabla F(x)$$

para um único $\mu \geq 0$ tal que $\| S(\mu) \| = \delta$, a menos que $\| S(0) \| \leq \delta$ e neste caso $S(0) = S^n$ é a solução. Para todo $\mu \geq 0$, $S(\mu)$ define uma direção de descida para F por x.

TEOREMA 3: Sejam $\| . \|$ e $|| . ||$ quaisquer normas no $\mathbb{R}^{n \times n}$. Então existem constantes positivas α e β tal que

$$\alpha \| A \| \leq || A || \leq \beta \| A \|$$

para todo $A \in \mathbb{R}^{n \times n}$. Em particular

$$n^{1/2}\| A \|_F \leq \| A \|_2 \leq \| A \|_F$$

e, para $p = 1$ ou $p = \infty$

$$n^{1/2}\| A \|_p \leq \| A \|_2 \leq n^{1/2} \| A \|_p .$$

A norma de Frobenius de A satisfaz

$$\| A \|_F = [Tr(A^T A)]^{1/2}$$

e, para todo $B \in \mathbb{R}^{n\times n}$,

$$\| AB \|_F \leq \min \{\| A \|_2 \| B \|_F, \| A \|_F \| B \|_2\}.$$

Portanto, para todo $v, w \in \mathbb{R}^n$,

$$\| Av \|_2 \leq \| A \|_F \| v \|_2$$

e

$$\| vw^T\|_F = \| vw^T\|_2 = \| v \|_2 \| w \|_2 .$$

Se A é não singular, então o operador norma induzido por $\| \cdot \|$ satisfaz

$$\| A^{-1}\| = \frac{1}{\min\limits_{v\in\mathbb{R}^n} \frac{\| Av \|}{\| v \|}} , \text{ com } v \neq 0.$$

TEOREMA 4: Seja $A \in \mathbb{R}^{n\times n}$ simétrica com autovalores $\lambda_1, \ldots, \lambda_n$. Então

$$\| A \| = \max \lambda_i , \quad 1 \leq i \leq n$$

Se A é não singular então

$$K_2(A) = \frac{\max \lambda_i}{\min \lambda_i} , \text{ com } 1 \leq i \leq n.$$

TEOREMA 5: Se $A \in \mathbb{R}^{nxn}$ tem autovalores λ_i, $i = 1,\ldots,n$, então $\lambda_i + \alpha$ é um autovalor de $A + \alpha I$, para todo real α.

BIBLIOGRAFIA

[1] J.M.Martínez, An Algorithm for Solving Sparse Nonlinear Least Squares Problems, Computing 39, 307-325 (1987).

[2] J.Dennis, R Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall Series in Comput. Math, Prentice-Hall, NJ. 1983.

[3] P.Gill, W.Murray, M.H.Wright, Practical Optimization, Academic Press, 1981.

[4] J.J.Moré, The Levenberg-Marquardt Algorithm: Implementation and Theory, Proc. Biennial Conf. Num. An., Dundee 1977, ed.by G. A. Watson. Lecture Notes in Math. Springer Verlag (1978).

[5] B.Noble, J.W.Daniel, Applied Linear Algebra, Prentice-Hall, Englewood Cliffs, NJ., 1977.

[6] D.Luenberger, Introduction to Linear and Nonlinear Progamming, Addison-Wesley, Massachusetts, 1973.

[7] R.S.Dembo, S.C.Eisenstat, T.Steihaug, Inexact Newton Methods, SIAM J. of Numer. Anal. 19 (1982), 400-408.

[8] R.S.Dembo, T.Steihaug, Truncated Newton Methods, Math. Programming 26 (1983), 190-212.

[9] J.J.Moré, B.S.Garbow, K.E.Hillstrom, Testing Unconstrained Optimization Software, ACM Transactions on Mathematical Software, Vol. 7, Mar 1981, 17-41.